EXPRESS MAIL LABEL NO.: EL660499973US
Attorney Docket No.:1378-1-1
Customer ID 25207

APPLICATION FOR UNITED STATES PATENT

METHOD AND SYSTEM FOR PROCESSING AND AGGREGATING MEDICAL INFORMATION FOR COMPARATIVE AND STATISTICAL ANALYSIS

INVENTORS:
HARRIS BERGMAN
DAVID KU
BRIAN NIEMEYER

CROSS REFERENCE TO RELATED APPLICATION

This application claims benefit of copending U.S. provisional application No. 60/242,028, filed October 20, 2000, entitled "Method and system for processing and aggregating medical information for comparative and statistical analysis", the disclosure of which is incorporated in its entirety herein by reference.

FIELD OF THE INVENTION

The present invention relates to a method and system for processing and aggregating medical information for analysis, including distributing the initial data acquisition among multiple medical practices, transferring data over the internet, storing data in a centralized database, and providing internet-based applications and services using the data—aggregated or individually—to be used in the care or management of patients.

BACKGROUND OF THE INVENTION

Few medical tests are 100% accurate. Even with the best data made available to the physician, medical errors still occur. Recently, several decision support tools—often embodied as software—have been developed to address the problem of misdiagnosis. For example, a pharmacist's label-printing software may connect to software that checks for drug interactions. These systems have several limitations that have hindered their adoption and reduced their benefit to the general public. Two major limitations are that they are often based on small clinical studies and that their use adds significant work and/or time for the physician.

Computer-aided detection and/or diagnosis ("CAD") is a class of systems that analyze medical data to help a physician determine a diagnosis. In the field of radiology, CAD systems have been developed to look for abnormalities in chest x-rays, heart scans, mammograms, and the like. They work by performing image processing on digitized radiological examinations (both native digital and digitally-scanned film), identifying

potential abnormalities and measuring their visual properties, and then determining whether these properties are indicative of a positive finding. The determination process may involve the use of empirical equations, rules (i.e., an "expert system"), or artificial intelligence; in any case, the specific parameters and weights used in this process are based on the results of clinical studies of patients. As with most scientific models, the accuracy of the CAD system is related to the sample size of the group (in this case number of patients) on which it was developed.

The only known presently commercially-available CAD system for the detection of breast cancer suffers from both the limitation of having been developed on a small sample of patients and also adding significant time that the physician must spend to interpret the mammogram and use the system. This system is a stand-alone computer-device mounted to a film reading station. Film is inserted into the device and, several minutes later, an analysis is presented. While it is processing the physician must wait. The decrease in physician productivity hinders the acceptance and use of these systems.

The present invention solves these and other limitations. The method and system described herein provides the infrastructure by which effective data-driven applications such as medical decision support or epidemiology research can be performed with high accuracy, ease-of-use, and portability.

Other objects, features, and advantages of the present invention will become apparent upon reading the following detailed description of embodiments of the invention, when taken in conjunction with the appended claims.

BRIEF DESCRIPTION OF THE DRAWINGS

The invention is illustrated in the drawings in which like reference characters designate the same or similar parts throughout the several figures of which:

Fig. 1 is a object model view of the system architecture of a preferred embodiment of the present invention.

Fig. 2 is a schematic view of the integrator.

Fig. 3 is a schematic view of the CAD preprocessor.

Fig. 4 is a schematic view of the exam flow overview.

Fig. 5 is a schematic view of the physician website map.

Fig. 6 is a schematic pictorial view of the system.

Figs. 7A and 7B are screen shots of an overview of the system.

Figs. 8A-8W are additional presentation views of aspects of the invention.

DESCRIPTION OF THE PREFERRED EMBODIMENT

The motivation for inventing this method and system is to make a product that encourages physicians to send patient information to the central database. As the database grows, many valuable applications that require aggregated medical information can be deployed, such as CAD.

In a preferred embodiment of the present invention the system comprises three parts: 1) a client module ("client") in a doctor's office, 2) the central system of a database and connected servers, loaders, and unloaders, and 3) at least one web-browser running at least one application.

The client in the doctor's office initially obtains the medical information. Depending on the type of information to be transmitted, the client could take several forms, such as, but not limited to, a web-browser, medical device, film digitizer or other form known to those skilled in the art or developed hereafter. Regardless of its form, the client will perform certain tasks: acquire medical information in digital form, perform some processing of the medical data, be attached to the Internet, periodically initiate a secure and/or encrypted connection between itself to the central database / server / loader over the Internet, and transmit the medical information across the connection.

The central system consists of the database and connected servers, data loaders, and data unloaders. The central system may be behind a firewall, Virtual Private Network or other device. Servers form connections to the clients mentioned above. At least one data loader takes the medical information deposited on the server and loads the data onto the appropriate tables in the database. At least one application server can query the database to perform analyses of the medical information on individual or aggregate (personal identifiers redacted) basis, for part of an Internet-based application. Analyzed data can also be stored on the database. The data unloaders and servers act as the intermediary between the database and the application.

Web-browsers can access applications that utilize the analyzed data from the database. Confidentiality of patients' data can be maintained by using encryption or similar technology over secure network. Applications can be developed for physicians, patients, or third parties. Potential applications range from patient registration in a doctor's office to the real-time comparison of a patient's chest x-ray to those of thousands of other patients.

CAD is one application that is well suited for the system of the present invention. In this case, patient information includes test data such as radiological images, a radiological report (i.e., the interpretation), and possibly additional, confirming reports (e.g., a pathology report, surgical notes, and the like). The CAD application would compare and analyze a new patient's test data against the aggregated data to suggest an interpretation. From a web browser, a physician can access the CAD results and make a more accurate diagnosis. Because the development and updating of CAD applications require both raw test data and confirmed diagnoses, the system can also extend an application to permit physicians to add confirmed results to a patient's record when the results become available. Each patient added to the database, whose record contains both raw test data and the corresponding confirmed results, can then be used to update the CAD application.

The key to the system is to "close the results loop," that is, to obtain not only patient test data but also the confirmed results. Critical to the commercial success of such a system is the development of a broad variety of tools, harnessing the system, to improve the productivity and quality of a physician. Web-based applications such as

report composition and patient registration provide incentives to add their patient information to the central database. Additionally, applications that increase the productivity of nurses and technicians can also be incorporated. An advantage of such applications is to encourage the entire staff of a medical practice to keep data stored in the central system. In doing so, additional information can be gathered into the database. An example of this kind of application is an online patient registration service, whereby patients type in their medical histories (for example), so that a nurse does not have to do so later.

The first application is directed to the detection of breast cancer in mammograms. The application streamlines the generation of the mammography report, and organizes and transports the medical images and reports in an efficient manner over the Internet to referring doctors, patients, and care providers. In a preferred embodiment, the application and underlying system utilize the newly available Internet as a Wide-Area-Network with high bandwidth, which was not part of healthcare information technology (HIT) solutions just a few years ago. All reports and images are available anytime, anywhere. A radiology practice also has the opportunity to get inside their patients' homes through "active letterhead" co-branding; patients can easily learn about their mammographer and other services the practice provides.

Behind the scenes, the system's application server compiles and archives patient data. With its permanent archive, the system can serve as a fulfillment center for distributing patient information. The database may be mined for reports on which patient sector best benefits from more frequent scans. Likewise, the cost-benefit analysis can be

made for less frequent scanning of younger women. For example, data from the archive can be sold to insurance companies to study outcomes and quality control. The mammography CAD interpretation system can become more accurate as the patient archive grows in size and the CAD is updated on an increasingly more robust patient population.

In cases where the image data is small and can be transmitted quickly to the central server, it may not be necessary to have parts (1) and (2) on the Integrator; instead, all the parts can be performed at the Central system.

Page 5 makes it seem like the images are sent to the central system and all analysis occurs there. In fact, the Integrator (residing at the hospital) does the initial processing before sending the images to the central system.

This is necessary for mammograms because the images are very large (over 100 MB per patient!); for other types of exams, this is not so necessary.

Computer-Aided Diagnosis (CAD) was invented to help radiologists make more accurate diagnoses. These systems can make an objective "second opinion," with which the radiologist can use. CAD algorithms in the field of radiology typically have three parts:

1) Feature extraction, where abnormalities of interest are isolated from the rest of the image. Extraction involves image processing techniques.

2) Feature analysis, where visual properties (such as size, darkness, border shape, etc.) of the extracted abnormality are measured.

3) Computation of the result. A "diagnosis" is calculated from the properties measured in (2). The relationship between the properties and the diagnosis is often very complex. Expert systems (e.g., "rules") and artificial intelligence (e.g., "neural networks" or "Bayesian networks") have been used to determine the relationships. Typically, the relationships are empirically determined and can be made more accurate when there is a large amount of validated data (feature properties and a corresponding confirmed diagnosis) from which to determine the relationships.

The results from (3) can be presented to physicians in many fashions, from a paper note indicating the result to an annotated digital image.

The present invention provides a system for performing CAD. In the invention, parts (1) and (2) are performed on the Integrator, and part (3) is performed at the central system. The reason for doing so is that it takes a good deal of time to transmit the images from the hospital to the central system. By performing the extraction and analysis steps at the hospital, the diagnosis can be received at the central system in a most expedient manner.

For purposes of displaying the results to the physician in a friendly manner, the Integrator also generates small (less than 100KB) versions of the large images and transmits them with the feature analysis data. Thus, the computed result, or diagnosis, can be visually displayed with the small version of the image; using a Web server at the central system, physicians can get access to the results from a Web browser.

The above describes a CAD system for radiological images. Other types of medical analysis can also be performed with the present invention. Other web applications tied into a CAD service can gather patient information such as current medications and family history. The broad medical data can also be analyzed in a manner similar to step (3) above, for determining things like drug interactions and risk factors for diseases.

In summary, the value of this system is in having an infrastructure by which physicians send and store medical information on a central database, so that the database grows at a fast rate and can support applications that analyze aggregated medical information. Furthermore, the present invention can protect intellectual property and confidentiality of CAD software and results, and perform CAD using a real-time database in an expeditious manner.

Further aspects of the invention and a systems architecture overview are shown in the following section having the heading "Systems Architecture Overview."

4. Systems Architecture Overview

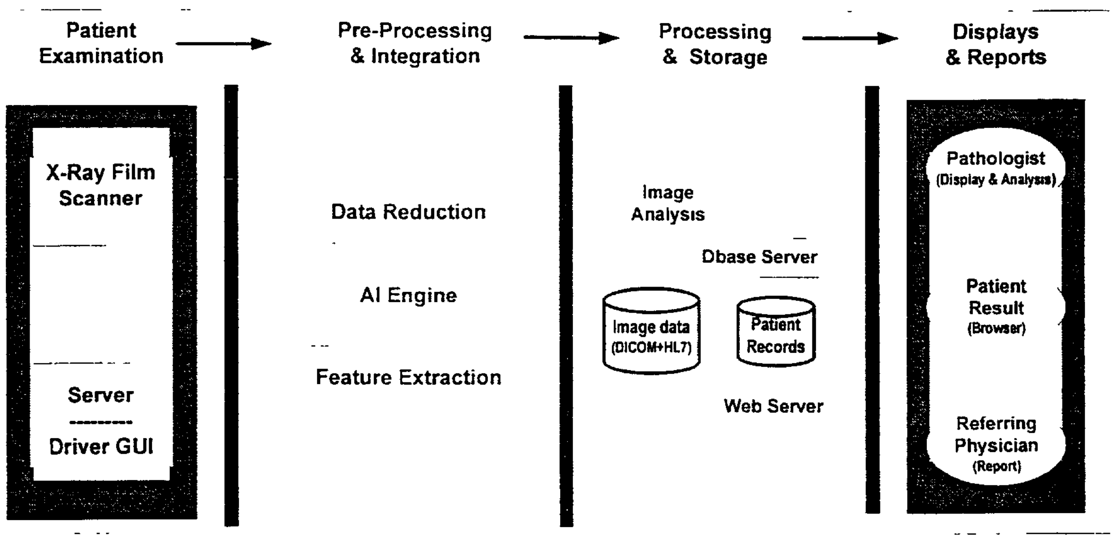

Figure-1: Architecture Overview Diagram No. 1

<u>The first logical step in the above architecture is the Patient Examination.</u>
Patient's records and/or demographic data are collected at this station. Then, after examination, the x-ray films are scanned, digitized and stored as four (4) different images. It is noted that each X-ray examination produces four (4) x-ray films for each patient.

<u>The second logical step of the architecture is the preprocessing and integration.</u>
Here, the system performs data manipulations, such as data reduction and data compression, on the stored digitized images. The system uses an Artificial Intelligence (AI) engine for feature extraction of the images. The system generates separate image data types from the originals.

1. Reduced data file — approx. 100-Kb size.
2. Full Resolution DICOM image file — approx. 150-MB size.
3. JPEG Image (4 Thumbnails) — (1 for each film/image) approx. 150-Mb size.

<u>The third logical step of the architecture is the processing and storage.</u>
This process involves the Database Engine which provides services such as database loader, records, objects, security, image handling, logging, HL7 (proprietary),
Also, the step includes a Web Server, which is responsible for providing results.
The Radiologist reviews the analysis and associate images provided by **SmartMamm**^{SM}.

<u>The fourth logical step of the architecture is the displays and reports.</u>
The process provides enhanced displays to the Radiologist using JPEG format with a client browser interface. Currently, the resolution is 256 x 256 pixels.
Also, the patient can access her results (JPEG) from a home computer using her client browser. The resolution here is also 256 x 256 pixels. Data results are displayed in readable format.

5. Three Architectural Models

MediZeus's business proposition involves the installation of turnkey systems at different types of medical facilities. This document provides technological implementation that supports all the following types of possible MediZeus's client's facilities.

A. Remote Clinic Model:

This is a standalone model where the Radiologist operates in a clinic facility. The facility could be in a medical center or in a local, private building. The patient usually initiates the examination request and it is performed at the nearby clinic. MediZeus system will reside in a standalone mode at such facilities. All data inputs, computations, analysis and results will be performed at that location. Image data (reduced, DICOM, JPEG and others) will be sent to the MediZeus Data Center via a DVD-ROM Compact Disk. Web Service for SmartMamm patients will be located at the MediZeus Center.

B. Stand-alone, Non-Networked Model

With this model, the mammography examinations are conducted as part of a hospital facility. Patients obtain their x-ray examinations at one or many x-ray operations within the hospital system. In this environment, the MediZeus system must interface with the hospital admission database to extract patient or demographic data. All data inputs, computations, analysis and results will be performed at that location. Image data (reduced, DICOM, JPEG and others) will be sent to the MediZeus Data Center via a high-speed, private wide-area network link. Web Service for SmartMamm patients will be located at the MediZeus Center.

C. Wide Area, Networked Model .

A main data center operation will serve as data repository for the large patient imaging data. Image data is estimated at 150-Mb file per patient with an annual estimated population of over 40 million examinations. 1 intermediate-term storage requirements for this data repository as over 5-PB (Pendabytes). This repository will contain patients' images, personal records, demographics and analysis data

All scanning, digitization, reduction will be acquired at the clinics. The back-end processing, data base and AI engines, storage and Web server functions will occur at the MediZeus Data Center. The analysis produced by the process will then be sent back to the requesting clinic for display and for additional data input. Web Service for SmartMamm patients will be located at the MediZeus Center.

This model requires a powerful, processing architecture to handle the processing, the comparison of newly scanned images with previous images. It will also handle the comparison of previous analysis and demographic data with new analysis.

This document considers the Standalone, non-networking Model in a Hospital/Clinic. The other consideration is the architecture comprising of both the Standalone, non-networking Model together with the resources at the MediZeus Center

IBM believes that MediZeus must be able to get this product, **SmartMamm**SM, into the marketplace as soon as possible. Therefore, we included this considerations into our architectures to facilitate both

- **Phase-1: Quick Deployment of modified, current technology**
  High performing, redundant ı servers, software, local/online storage, tape backups, web access and security for **SmartMamm**SM clients.

- **Phase-2. Later deployment of Enhanced, Robust, high-performing technology.**
  Highly available, redundant servers, software, display technologies, centralized online storage, online tape backups, personalized web access, security and authentication for **SmartMamm**SM clients.

6. Phase-1: Roadmap Architectures

We present the following scenarios, representing different architectures for providing Smartmamm technology with solutions. MediZeus has a stated preference for the **Linux** operating system.

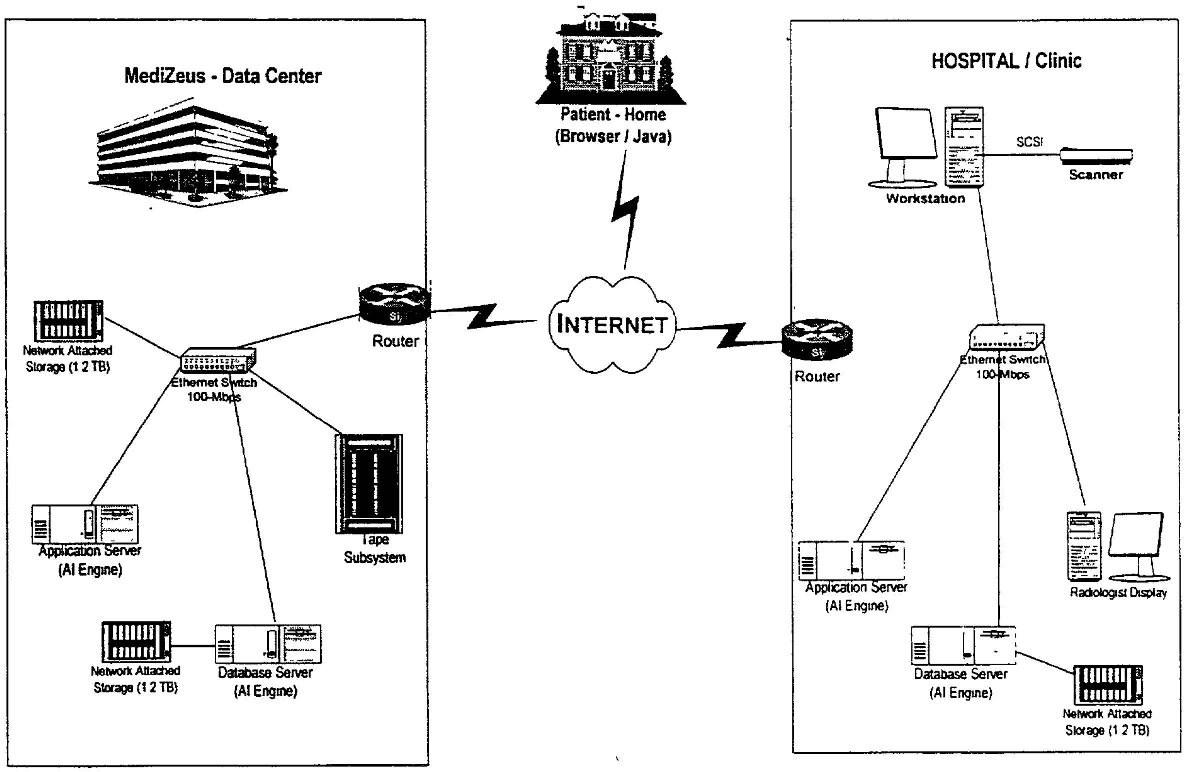

Figure-2: Logical Overview of a Stand-Alone Architecture

6a. Standalone, non-network Model:

Definition:
We define this model as a complete, self-contained system. All data inputs, manipulations, analysis and display activities are performed at the predetermined radiologist environment. The radiologist may be in a hospital environment or a stand-alone clinic facility.

This model is described as a non-networking architecture because it does not require the transfers of data inputs to remote Data Center to complete the analysis of the mammography exams.
All data, reduced, DICOM, JPEG and others will be transferred to the MediZeus's main Data Center at a later time. This might probably be a post-analysis activity.

Advantages of this model include:
- Faster response times. All computer processing power required will be local.
- Overall system costs. The network cost in this model is lesser since real-time transfers are not required to complete analysis.

- Scaling: As demands for more computer processing or storage grows, MediZeus can meet these needs by deploying additional hardware and/or storage or enhancements without a radical change in the system architecture.

The Data Center at MediZeus will comprise of similar systems architecture with more data storage than the Hospital/Clinic models. As data storage requirements grow at the main Data Center, the architecture can evolve into either a fully scaled Storage Area Networks (SAN) within MediZeus or as a Managed Storage Services offering .

Managed Storage Services offer dynamic storage solutions based on the premise of "Pay-As-You-Go", where storage becomes available on demand. This is an alternate solution to having MediZeus hosting the very, large storage equipment on MediZeus's premises along with qualified resources to manage and administer the storage subsystems.

Phase-1 Hardware Components

This is an architectural estimation of technology to support the baseline configuration of the standalone, non-networked model. The components described below are not suggested configurations, sizing, pricing, etc.

A. Scan Workstation(s): (with monitors and keyboard, mouse, etc.)

Each Scan Workstation will comprise of.
•Intel-based PC with a SCSI adapter connecting to the external Scanner device.
•Disk capacity for a **temporary storage** buffer (2 x 150-Mb per scan).
•High-speed Ethernet connection to the Ethernet Switch (100-Mbps or 1-Gigabit).

Recommended Software:
•Scanner device with appropriate Scanner Drivers and GUI software
Recommended Operating System: (one OR the other, not both)
•Microsoft NT Operating System (if the current Scanner software only runs on Microsoft)
•Citrix MetaFrame Interface to a LINUX Server (if the current Scanner software can execute within a Citrix Frame environment).

Image Data Storage:
X-ray films that are scanned and digitized will be stored on the Server, not on the Scanning workstation. Large buffers will be allocated on the Server disk system for storing scanned images for data reduction purposes.

Advantages to this architecture include:
1. Data integrity: The image data will reside at only one location-the Server.
2. Processing Time: This approach eliminates the need for file transfers of scanned images from the Scan Workstation to the Server. The Server can immediately work on its internal buffers.
3. Scaling: Since there may be many Servers, it is feasible to spread the internal buffers across multiple Servers for performance.

B. Application Server(s)

Each Application Server will comprise of: (

- •Intel-based, High Performance machine (with monitor(s), keyboard(s), etc.)
- •High Memory at least 2-GB RAM with L2-Cache
- •Disk capacity for a **storage** (20 x 150-Mb per scan) for 1 year (3.6-TB) (based on: 20-days/month, 20 scans per day, 240 days/year).
- •High-speed Ethernet connection to the Ethernet Switch (100-Mbps or 1-Gigabit)
- •A RAID-5 technology (for disk failure protection),

Recommended Software:

- •MediZeus's AI (Artificial Intelligence) Engine (C++)
- •MediZeus's Data Reduction, Extraction modules (C++)
- •Other MediZeus software related to data manipulation
- •Systems and Storage Management Software
- •Security Management Software

Recommended Operating System:

- •LINUX

Scaling of the Application Servers:
Initially, there will be a minimum of two (2) Servers to balance workloads and provide redundancy. As the operation grows, the Servers at this level of operation, will become a high-performing Cluster of LINUX Servers, with high-computing capacity.

C. Network Attached Storage (NAS)

Network Attached Storage devices provides storage for images, data, database files and other Smartmamm file. It uses TCP/IP protocol to communicate between the servers on the LAN.

Initial configuration of each of the NAS devices should be 1.2-TB disk capacity. Additional NAS devices can easily be installed in the network for expansion.

As the image data grows, MediZeus can upgrade the data Storage solutions to Storage Area Network (SAN) devices to support terabytes of data. The projected storage requirements for the Smartmamm marketplace is approx. 5-PB.

D. Network Attached Storage (NAS)

A magnetic, Tape Library Subsystem will provide backup for critical patients data and images and other database files. Additional robotics arms and library devices can be added for expansion, speed of operation and performance purposes.

E: Ethernet Switches

High-speed Ethernet Switch with at least 4 x 1-GB Ethernet ports for Server connections
Scan Workstations, Servers, Display Servers, Firewalls and other internetworking devices will connect via the Ethernet Switch(es).

F. Firewalls with Intrusion Detection Device(s)

The Firewall(s) will protect proprietary patient's records, images and data from other networks within a hospital or clinic complex. The Intrusion Detection Scheme provides additional layer of security from Denial Of Services (DOS) attacks from either the Internet or within a complex.

G. Web Server

A Web Server, (, an Apache-based appliance, will service the web requests from patients' home computers. Access to this web server is through the Internet. It will cache static MediZeus pages and relevant images. It will cache result data and images that are appropriate for remote viewing.
The Web services will support industry Web Browsers, such as Internet Explorer and Netscape Communicator.

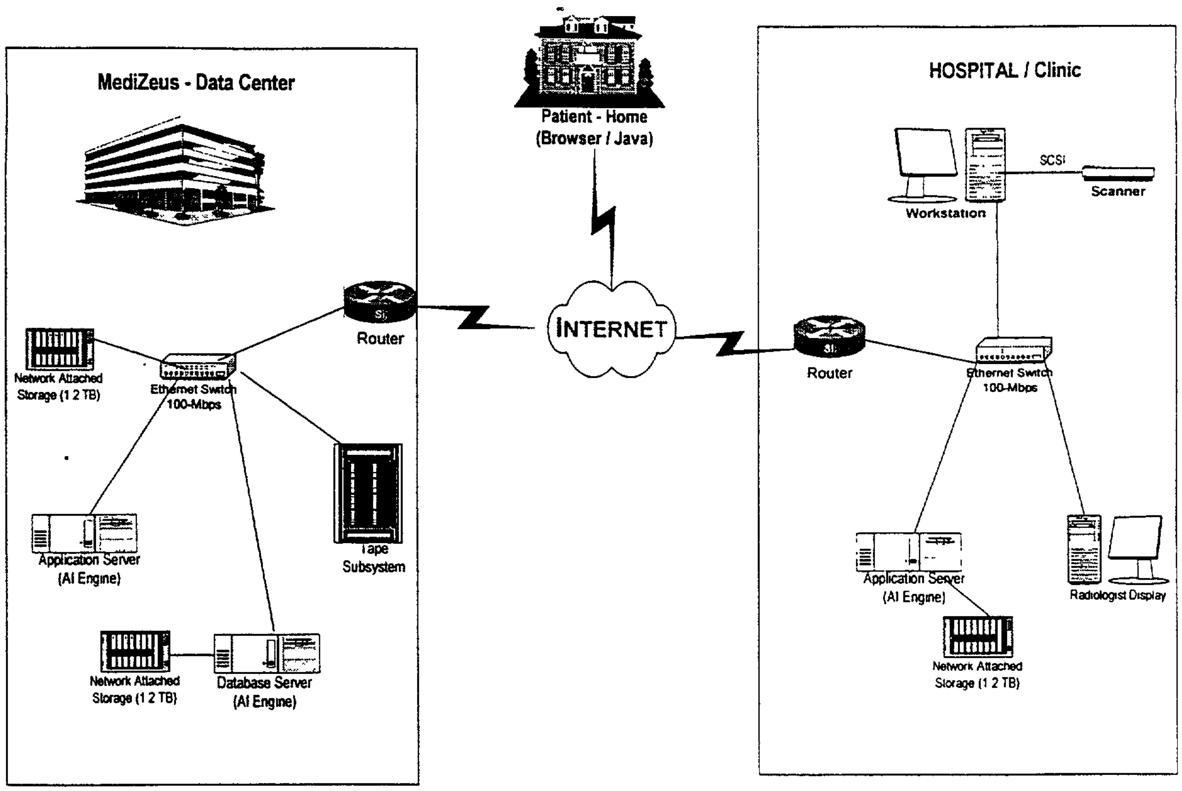

Figure-3: Logical Overview of a Network Architecture

6b. Wide Area, Network Model:

Definition:
We define this model as a complete, remote and local system. All data inputs, manipulations, analysis and display activities are performed at the predetermined radiologist environment. The radiologist may be in a hospital environment or a stand-alone clinic facility.

This model is described as a networked architecture because it requires the transfers of data inputs to remote Data Center to complete the analysis of the mammography exams. All data (reduced, DICOM, JPEG and others) will be transferred to the MediZeus's main Data Center, immediately for processing and the results sent back to the requesting facility for displays and reports.

Advantages of this model include:

- Data Integrity: Patient's reduced images, data, analysis reside at one central location
- Computing: The computing systems (and Engines) will be more powerful at the centralized location.
- Comparison: Existing data comparison and other correlation activities can be performed quickly and extensively at the central location.
- Economy: Since processing power is centralized, there is reduced need for high-end computing at the different medical facilities.

Disadvantages of this model include:

- Response times. Data must be transferred first, processing and results sent all over the wide area network. There will be network delay introduced into the flow of the analysis.
- Network: Network outages, congestion, unscheduled downtime will have impact on the overall availability of the operation.
- Cost: High-speed, redundant network links can introduce very high, recurring costs to the operation. Depending on the Response Time requirements for analysis to be completed, transfers of 150-Mb files across wide can require very high speed links.

Although only a few exemplary embodiments of this invention have been described in detail above, those skilled in the art will readily appreciate that many modifications are possible in the exemplary embodiments without materially departing from the novel teachings and advantages of this invention. Accordingly, all such modifications are intended to be included within the scope of this invention as defined in the following claims. It should further be noted that any patents, applications or publications referred to herein are incorporated by reference in their entirety.